

## Durham Research Online

---

### Deposited in DRO:

28 September 2018

### Version of attached file:

Published Version

### Peer-review status of attached file:

Peer-reviewed

### Citation for published item:

Bower, EdwardK. M. and Cooper, LaurieP. and Roberts, GarethA. and White, JohnH. and Luyten, Yvette and Morgan, RichardD. and Dryden, DavidT. F. (2018) 'A model for the evolution of prokaryotic DNA restriction-modification systems based upon the structural malleability of Type I restriction-modification enzymes.', *Nucleic acids research.*, 46 (17). pp. 9067-9080.

### Further information on publisher's website:

<https://doi.org/10.1093/nar/gky760>

### Publisher's copyright statement:

© The Author(s) 2018. Published by Oxford University Press on behalf of Nucleic Acids Research. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

### Additional information:

## Use policy

---

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

# A model for the evolution of prokaryotic DNA restriction-modification systems based upon the structural malleability of Type I restriction-modification enzymes

Edward K.M. Bower<sup>1</sup>, Laurie P. Cooper<sup>1</sup>, Gareth A. Roberts<sup>1</sup>, John H. White<sup>1</sup>, Yvette Luyten<sup>2</sup>, Richard D. Morgan<sup>2</sup> and David T.F. Dryden<sup>3,\*</sup>

<sup>1</sup>EaStCHEM School of Chemistry, University of Edinburgh, The King's Buildings, Edinburgh EH9 3FJ, UK, <sup>2</sup>New England Biolabs, 240 County Road, Ipswich, MA 01938-2723, USA and <sup>3</sup>Department of Biosciences, Durham University, South Road, Durham DH1 3LE, UK

Received October 05, 2017; Revised August 07, 2018; Editorial Decision August 08, 2018; Accepted August 21, 2018

## ABSTRACT

**Restriction Modification (RM) systems prevent the invasion of foreign genetic material into bacterial cells by restriction and protect the host's genetic material by methylation. They are therefore important in maintaining the integrity of the host genome. RM systems are currently classified into four types (I to IV) on the basis of differences in composition, target recognition, cofactors and the manner in which they cleave DNA. Comparing the structures of the different types, similarities can be observed suggesting an evolutionary link between these different types. This work describes the 'deconstruction' of a large Type I RM enzyme into forms structurally similar to smaller Type II RM enzymes in an effort to elucidate the pathway taken by Nature to form these different RM enzymes. Based upon the ability to engineer new enzymes from the Type I 'scaffold', an evolutionary pathway and the evolutionary pressures required to move along the pathway from Type I RM systems to Type II RM systems are proposed. Experiments to test the evolutionary model are discussed.**

## INTRODUCTION

Prokaryotic restriction-modification (RM) systems provide a major defence against invading foreign DNA (1–4) and as such their genes are found in over 96% of bacterial genomes and over 99% of archaeal genomes (5,6). A typical RM system (7–12) includes a restriction endonuclease (REase), whose cleavage of DNA is triggered by the recognition of a specific DNA sequence on foreign DNA. The other

constituent part of the RM system is a methyltransferase (MTase), whose action prevents cleavage of host DNA by methylating the target DNA sequence. Given their significant role in protecting the host cell, it is surprising that RM systems are not essential to prokaryotic life. As such, RM systems should be viewed as necessary for the survival of the population, and not the individual cell; RM activity is the main method to prevent the spread of foreign DNA in a population (1,3,11,13–16) although additional roles have been proposed (3).

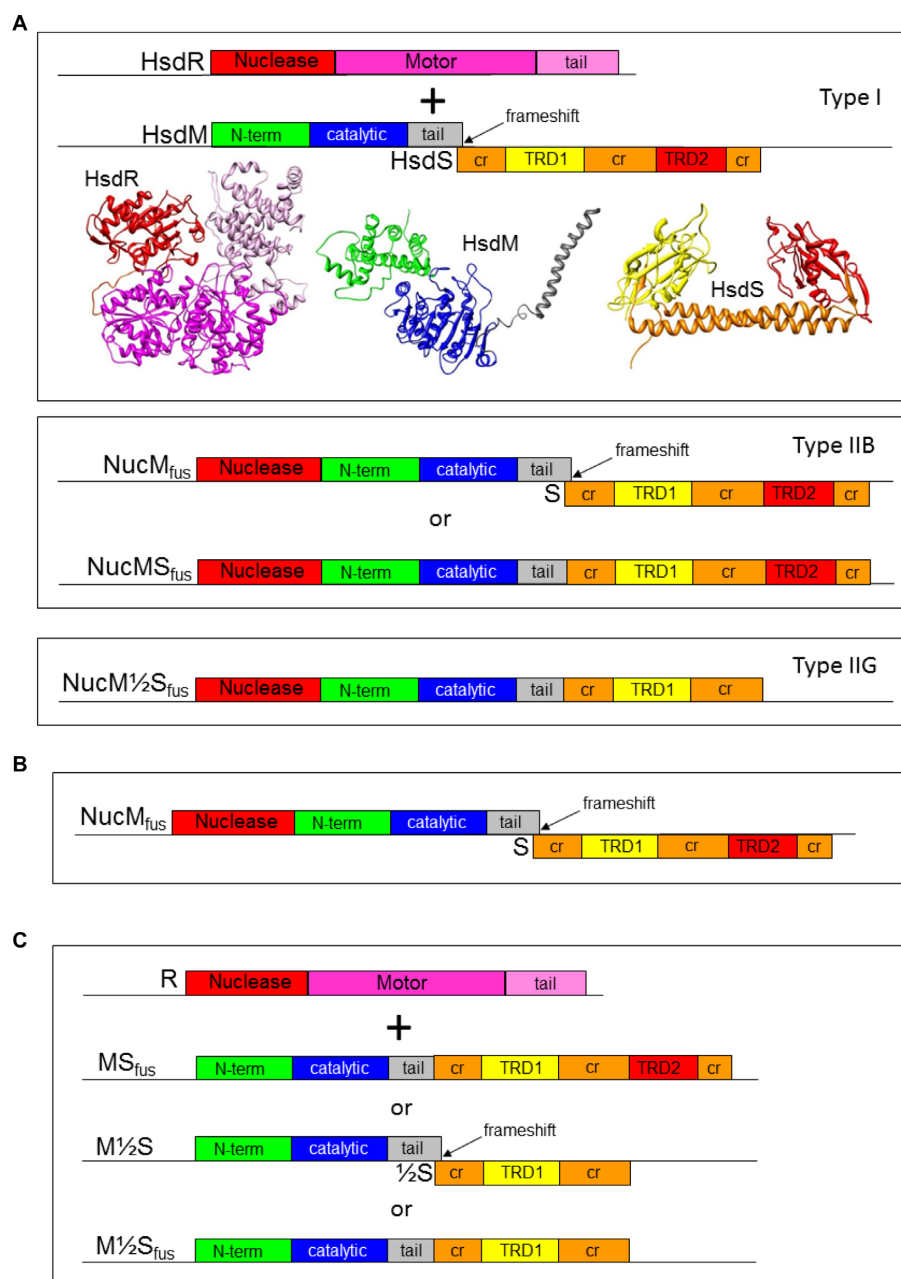
In some cases, RM functions are carried out by separate REase and MTase enzymes. However, in many systems both of these activities are fulfilled by a multi-subunit protein or even a single polypeptide (7,17). Hence, the RM systems show great variety in protein structure and gene sequence. To date, there are three classes of RM systems (Types I to III) and one class operating only on methylated DNA and thus lacking the modification function while retaining the restriction function (Type IV). These Types are separated due to differences in composition, target recognition, cofactors and the manner in which they cleave DNA (18). The defining characteristic of Type II RM systems, and perhaps the most important in terms of their use to molecular biology, is that their REase cleaves double stranded DNA at fixed, easily identified positions at or near to the target sequence (19).

## Type I RM enzymes and their structural malleability

Type I systems were the first RM systems to be discovered (7,8,19). They are large hetero-oligomeric complexes, which perform cleavage of DNA away from their recognition site, in an ATP-dependent reaction (8,20–24), Figure 1A. A Type I restriction enzyme is composed of three separate sub-

\*To whom correspondence should be addressed. Tel: +44 191 3341200; Email: david.t.dryden@durham.ac.uk

Present address: Edward K.M. Bower, Structural Genomics Consortium, University of Oxford, Old Road Campus Research Building, Roosevelt Drive, Oxford, OX3 7DQ



**Figure 1.** (A) A cartoon comparison of the domain structures encoded by Type I, Type IIB and IIG RM systems. The HsdR subunit contains three key domains, the N terminal nuclease (red), and the motor domain (dark pink) and tail region (light pink). The HsdM subunit contains three domains, the N terminal (green), the catalytic (blue), and a tail region (grey). The HsdS subunit is composed of two target recognition domains (TRDs, yellow and red), which are linked by regions of conserved sequence ('CR', orange). Structures derived by X-ray crystallography and modelling (24), are shown with the same domain-specific colouring (not to scale). The same colour scheme is used for the domains in Type IIB and IIG RM systems. (B) In this work, the fusion of the Nuclease domain of HsdR to the N-terminus of the HsdM is investigated. (C) Further domain re-arrangements between the HsdM and HsdS are also investigated in this work.

units. These subunits are denoted by Hsd (host specificity for DNA) R for the restriction subunit (~130 kDa), M for the MTase subunit (~60 kDa), and S for the sequence-recognition specificity subunit (~50 kDa). The ~440 kDa restriction complex has a  $R_2M_2S_1$  stoichiometry, whilst a  $M_2S_1$  stoichiometry acts as a cognate MTase for the system. Type I enzymes use energy from ATP hydrolysis to translocate DNA. The HsdR subunit binds both ATP and  $Mg^{2+}$  in order to perform the complicated process involved

in producing double strand breaks in unmethylated DNA. The Type I enzyme binds its recognition sequence and the motor domains in the HsdR reel the DNA in towards the enzyme and cutting occurs when two HsdR motors collide. This can occur at anything from 40 bp to many kb away from the recognition site but is generally about half way between one site and the next target site.

Specific DNA sequences are recognised by the HsdS subunit (22–27). For the most part, these sequences pos-

sess the same general organisation of three specific nucleotides followed by a variable spacer of five to eight non-specific nucleotides (N), and then a further three to four specific nucleotides. For example, the target sequence of SauSTORF499P from *Staphylococcus aureus* clonal complex CC398 is ACC(N)<sub>5</sub>RTGA (26). The S subunit contains two target recognition domains (TRDs), separated by a central domain, which is conserved in members of the same family (20,25–27). The N-terminal TRD is specific for the 5' part of the bipartite DNA sequence, while the C-terminal TRD recognises the 3' part. The central conserved domain serves to coordinate interactions with the other subunits and more importantly, to separate the TRDs to a defined distance matching the length of the non-specific DNA spacer in the target sequence (20). The presence of additional N terminal and C terminal conserved domains, which together match the sequence of the central conserved domain, indicated a circular arrangement of the structure (28–32) which was confirmed by crystallography (33–35).

Alterations to the TRDs, which are encoded by the *hsdS* gene, can establish a new DNA specificity. TRDs may be swapped, truncated, switched and the number of non-specific nucleotides can be increased by extending the conserved sequence separating the two TRDs to engineer new specificity (25–27,36–40). Not only do the *hsdM* and *hsdS* genes share the same promoter but their open reading frames overlap at the junction created by the end of *hsdM* and the start of *hsdS* (41). Hence, the subsequent translation is coupled and during translation, a jump is required to create the two separate polypeptides. Roberts *et al.* were able to remove this frameshift from the MTase genes of the EcoKI Type I enzyme to create a fusion of the M and S subunits (42). This protein product showed full RM activity *in vivo* and was also successfully over-expressed and purified. With the addition of stoichiometric amounts of EcoKI HsdM protein, the purified fusion formed an active restriction complex *in vitro*.

## Type II RM enzymes and their classification

The most well-known and commonly used Type II REases fall into the subcategory of Type IIP as they bind and cleave palindromic sequences. The Type IIP REases, such as EcoRI and BglII, are 'Orthodox' ~60 kDa homodimeric complexes, which cleave within or next to their recognition sequence (12,18). The associated MTases are usually thought of as monomers of ~30 kDa although at least some function as dimers (17,43,44). Not all of the Type II RM systems conform to the narrow definition of Type IIP systems, and so they are separated into other sub-categories (18), Figure 1A. The factors that differentiate the many Type II sub-types are: the nature of the recognition sequence, tertiary and quaternary structure, and the type of cut produced on DNA. Type IIB systems have a subunit organization and activity not unlike Type I systems although they lack the ATP-dependent DNA-translocating motors (45,46). They possess REase, MTase and two TRDs in one enzyme, and can methylate either symmetric or asymmetric sequences which are indistinguishable from the targets recognised by the Type I RM systems. They cleave DNA in a SAM-dependent reaction and do so either side of their

recognition sequence resulting in the removal of a short fragment. In effect they are 'motor-less' Type I RM systems. Like many Type IIB systems, IIG REases also encode REase, MTase and a single TRD in a single polypeptide. They are effectively half of a Type IIB RM enzyme or half of a 'motor-less' Type I RM system (47–49), Figure 1A.

## RM enzymes with similarity to Type I RM enzymes

In addition to the Type IIB and IIG RM systems discussed above, additional naturally-occurring, evolutionary intermediates between Type I and II systems are known. The Type ISP family possess all the functions of a Type I enzyme including the ATP-dependent motor, within a single polypeptide (SP) (50–52). The single TRD in a Type ISP system recognises a 6–7 bp asymmetrical target. They perform ATP-dependent dsDNA translocation and cleavage, and SAM-dependent modification. Type III RM enzymes are ATP-dependent hetero-oligomeric enzymes possessing REase and MTase in a single complex with a 'res' REase subunit and a 'mod' MTase subunit containing a well-defined TRD (9,53). The motor domain in the res subunit facilitates diffusion of the enzyme on the DNA rather than the directed translocation driven by the motor domain in the Type I HsdR subunits but the domains have the same evolutionary origin (53). There are also RM systems known colloquially as 'Type one and a half' systems comprised of a Type IIP REase and a Type I MTase (7) and more recently the structure of a Type IIP REase with a striking similarity to a Type I HsdS subunit has been determined (54).

## Evolution of RM enzymes

This work aims to investigate the hypothesis that Type I, Type II and Type III RM systems are evolutionarily linked. By making step-wise alterations to the subunits of the SauSTORF499P CC398-1 Type I RM system from *S. aureus* (15,25–27), active enzymes with novel specificities have been successfully engineered. Soluble fusions of the nuclease domain from HsdR to HsdM, HsdM to HsdS and HsdM to half-HsdS were produced. These new protein structures are comparable to Type II RM systems.

This leads us to propose a structural model for the evolution of RM systems which attempts to answer the following questions:

1. When did RM first appear?
2. Why did it appear?
3. What did the first RM system look like?
4. Why did it subsequently evolve to form the large range of RM variants observed today?
5. What was the evolutionary pathway?

## MATERIALS AND METHODS

### Chemicals, bacterial strains and phage

All chemicals were purchased from Sigma-Aldrich unless otherwise stated.

Bacterial strain *Escherichia coli* NM1261 ( $r_K^- m_K^+$ ) was used for assays for *in vivo* activity (27). Bacterial strains *E. coli* BL21 (DE3) ( $\Delta hsdS_B$ ) and DH5 $\alpha$  ( $hsdR17 r_K^- m_K^+$ )



were from New England Biolabs (Ipswich, MA). Bacterial strain *E. coli* ER2796, a kind gift from Dr E. Raleigh (New England Biolabs), was used in the SMRT sequencing protocol (55). *E. coli* ER2796 is a derivative of *E. coli* K12 with the phenotype  $\lambda$ -*fluA2*  $\Delta$ (*lacZ*)*r1* *glnV44* *mcr-62* *trp-31* *dcm-6* *zed-501::Tn10* *hisG1* *argG6* *rpsL104* *dam-16::Kan* *xyl-7* *mtlA2* *metB1* (*mcrB-hsdK-mrr*)114::IS10 lacking all host DNA methylation systems (55).

The pJF118His plasmid for expression of all genes has been previously described (27). Recombinant plasmids were isolated from transformed *E. coli* DH5 $\alpha$  cells and the desired DNA sequences were confirmed. The amino acid sequences of the HsdR, HsdM and HsdS used in this work are given in Supplementary Table S1.

Fusions of the DNA encoding the nuclease domain of HsdR to the 5' end of *hsdM* were created using PCR. The first RM fusion gene was created in a PCR using the oligonucleotides, 'Mu50nucuni TS' (5'-AGTCAGTCAGGGATCCAAGAAGGAGATATACATATGGCATAACCAAGTGAATACGC-3') and 'Mu50nucendalpha-CC398-1BS' (5'-CGTTGTTTTCAGTAATAGACATATTATCCCTGTCTCAGTCG-3'), with the template *hsdR* gene (SauN315ORF189P) previously ligated into pRSFDuet-1 (Novagen) (27). The second PCR was conducted using the oligonucleotides, 'Mu50nucendcoil-CC398-1TS' (5'-CGACTGAGACAGGGAATAATATGTCTATTACTGAAAAACAACG-3') and 'C398-1BS' (5'-GATCGAATTCGGATCCAATAAACATCTTTTGAAGTAATGAC-3'), with the wild-type CC398-1 MTase genes in pJF118His vector as template (27). Further *hsdR* to *hsdM* fusion constructs were created using the same outer primers ('Mu50nucuni TS' and 'CC398-1BS') and specific primers for the different regions of *hsdR*, to which the *hsdR* portion of the fusion would be truncated. The fusion constructs, the primers used to create them in PCR and their amino acid sequences are summarised in Supplementary Table S2. The fusion proteins retain the methionine encoded by the ATG codon at the start of *hsdM*. The gene sequences used as a source for this work are SauN315ORF189P, M.SauSTORF499P and S.SauSTORF499P from REBASE (6).

To create the CC398-1 MS fusion gene encoding the protein 'MS<sub>fus</sub>', the *hsdM* open reading frame was fused in frame to *hsdS* by the polymerase chain reaction (PCR) and the resulting product was ligated into the pJF118His vector. PCR using oligonucleotides 'HsdM-TS' (5'-GATCGATCGGATCCAAGAAGGAGATATACATATGTC-3') and 'MTasefusion-BS' (5'-GCACATTTTCTTTTGTGTATTACTCATCTCATCTTTCAACACCCCAAG-3'), with the wild-type CC398-1 MTase genes in pJF118His as template, generated a fragment comprising the 5' UTR upstream of *hsdM* and the entire *hsdM* ORF, fused in frame with the first 28 bases of *hsdS*. A PCR with a second pair of oligonucleotides, 'MTasefusion-TS' (5'-CTTGGGGTGTTGAAAGATGAGATGAGTAAATACACAAAAGAAAAATGTGC-3') and 'CC398-1BS' (5'-GATCGAATTCGGATCCAATAAACATCTTTGAAGTAATGAC-3'), with the wild-type CC398-1 MTase genes in pJF118His as template, generated a fragment comprising the last 29 bases of *hsdM* fused to the entire ORF of *hsdS*. These two PCR products were purified and fused

in a reaction primed with oligonucleotides 'HsdM-TS' and 'CC398-1BS'. The resulting product was purified and digested with BamHI. pJFMS was digested with BamHI, treated with Calf Intestinal Phosphatase and then ligated with the PCR product.

To create the 'Half S' MTase 'M<sub>1/2</sub>S', *hsdS* was truncated at the end of the central conserved region, directly before the start of the second TRD (equivalent to amino acid D220). PCR was performed with primers 'HsdM-TS' and 'CC398-1 TRD 1 BS2' (5'-GATCGAATTCGGATCCATCTTACCATTCTCATCTTTAAATCG-3') with wild-type CC398-1 MTase genes in pJF118His as template. The product of this reaction was subjected to agarose gel electrophoresis and the band of the expected size was excised, gel eluted, BamHI digested and then ligated into vector pJF118His. The *hsdM* to half *hsdS* fusion gene to produce the protein 'M<sub>1/2</sub>S<sub>fus</sub>' was made in the same way but with the fused MTase genes in pJF118His in the PCR.

The amino acid sequences of these constructs are given in Supplementary Table S3. The fusion proteins retain the methionine encoded by the ATG codon at the start of *hsdS*.

### Bacterial genome single molecule real-time (SMRT) sequencing

Non-methylating (*dam*<sup>-</sup> *dcm*<sup>-</sup>) *E. coli* ER2791 competent cells were transformed with a plasmid containing the target MTase and spread on a plate of lysogeny broth (LB) agar (25). Agar plates were supplemented with 100  $\mu$ g/ml carbenicillin, which acted as a selection marker for the expression construct. Plates were incubated at 37°C overnight. A colony of successful transformants was picked into 5 ml of LB supplemented with 100  $\mu$ g/ml carbenicillin and incubated overnight at 37°C whilst shaking. Cells from the subsequent culture were separated into 1 ml aliquots and harvested by centrifugation at 2380  $\times$  g for 15 minutes at 4°C. The Wizard Genomic DNA Purification Kit (Promega, Madison, WI, USA) was then used to lyse the cells and purify the genomic DNA. The quality of the genomic DNA preparations was initially assessed by agarose gel electrophoresis and from the shape of the absorbance profile from 240 to 340 nm. The DNA library for SMRT sequencing was prepared and subsequently analysed using a Pacific Biosciences sequencer as described in Anton *et al.* (55).

### Gene expression and purification of proteins

Overexpression of all genes was carried out in *E. coli* BL21 (DE3) competent cells, which were transformed with the plasmid expressing the target gene. Induction of expression was performed by adding IPTG to 1 mM and further growth at 20°C overnight (~18 h).

All proteins were expressed with a hexa-HisTag attached to the C-terminus of the HsdS part of the protein. After overexpression of target genes, the *E. coli* BL21 (DE3) cell pellets were resuspended in 20 mM sodium phosphate buffer with 500 mM NaCl (pH 7.5), 20 mM Imidazole and a dissolved EDTA-free protease inhibitor tablet (Roche), in a 1:10 (g: ml) ratio. The cells were then disrupted by sonication using a Soniprep 150 sonicator (Sanyo, Tokyo, Japan),

fitted with a 9mm diameter probe for ~20 minutes with intermittent cooling. Cells were then centrifuged at 7700 x g for ~45 min at 4°C. The supernatant was filtered through a filter unit (0.45 µm; Sartorius AG, Goettingen, Germany) and then applied to a pre-equilibrated Histrap FF 5 ml column (GE Healthcare) at a flow rate of 100 ml/hr. The flow-through was collected. The column was then washed with 100 ml 20 mM Imidazole buffer and the flow-through was collected. This was followed by an elution of the protein with ~10 ml of buffer supplemented with 500 mM imidazole, discarding the first 3 ml and collecting the next 6 ml. This was then concentrated to ~4 ml, using a 20 ml 30 000 MWCO Vivaspin concentrator (Sartorius).

The 500 mM imidazole-containing buffer was removed from the sample by buffer exchange. This was performed using a PD-10 desalting column (GE Healthcare). The protein sample was concentrated to 2.5 ml and loaded on to the PD-10 column equilibrated with 20 mM Tris-HCl pH 8, 10 mM MgCl<sub>2</sub>, 500 mM NaCl and 7 mM 2-mercaptoethanol buffer. After elution, sample concentration was determined by an A<sub>280</sub> reading. Part of the sample was immediately used in an assay to determine presence of DNA cleavage activity and the remainder concentrated in the Vivaspin concentrator (Sartorius). Samples were stored at -20°C after the addition of glycerol to 50% (v/v).

All HPLC analytical size exclusion runs were carried out using a BioSep-SEC-S 3000 (Phenomenex) column and a pH 6.5 buffer (20 mM Tris, 20 mM MES, 10 mM MgCl<sub>2</sub>, 200 mM NaCl, 0.1 mM EDTA, 7 mM 2-mercaptoethanol). This buffer was used to dilute the samples to a concentration of approximately 4 µM, 50 µl of which were then injected onto the HPLC system for each run. A flow rate of 0.5 ml/min was used for each run, which took approximately 10 min to complete. The absorbance at 280 nm was monitored and recorded by a data logger. The column was calibrated using several protein standards (Sigma Aldrich) of various concentrations, and a calibration curve was produced.

### Assessment of RM activity in vivo

The methods employed for assaying RM activity *in vivo* used the efficiency of plating (eop) of phage λ<sub>v</sub> prepared from either a strain lacking the MTase genes to obtain unmodified phage or a strain transformed with the MTase plasmid to obtain modified phage have been previously described (56,57). All assays were performed in triplicate either as spot tests (nuclease domain fusions) or as full plate tests (MS fusions). The spot tests are suitable when restriction is absent and the whole plate tests are required when restriction is present to better quantify the degree of restriction. The spot tests used for assaying the nuclease fusions often give a standard deviation of ~30% so a value of eop greater than one can occur (56). The promoter on the expression plasmids is slightly leaky so IPTG was not required to be added to the plates.

### In vitro DNA cleavage assay

Assays were conducted by incubating the enzyme under investigation with a library of plasmids. These plasmids were created by the ligation of known DNA sequences between EcoRI and BamHI sites of pUC19. The plasmids

in the library were based on the DNA sequence of phage PhiED1 (a gift from Garry Blakely, University of Edinburgh). These plasmids are numbered sequentially from 1E to 20E (omitting 3E and 8E). The plasmid and specific insert sequences of these plasmids have been described previously (27) and are given again in the Supplementary Table S4. Each plasmid contains a ~2.4 kb insert that was PCR amplified from *Bacteroides fragilis* phage PhiED1, ligated into vector pUC19. Collectively, the 18 plasmids contain >40 kb of known sequence and comparing their susceptibility to cleavage allows inference of REase specificity (27). MTases under investigation were supplemented with R subunit from *S. aureus* CC5 and incubated with the plasmid library in separate reactions. Reaction digests had a total volume of 50 µl and a typical digest was prepared using 5 µl of 10× NEBuffer 4 [New England Biolabs; 50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 1 mM dithiothreitol (pH 7.9)], 2 mM ATP, 0.64 mM S-adenosyl-L-methionine, 0.01 mg of bovine serum albumin and 10 µl of the enzyme stock. The enzyme stock was prepared in a volume of 50 µl with 5 µl of 10× NEBuffer 4 with final concentrations of 1.16 µM R subunit and 0.42 µM MTase, thus ensuring an excess of R over the MTase to give formation of the R<sub>2</sub>M<sub>2</sub>S<sub>1</sub> RM enzyme. Incubations were left for 12 min in a water bath set at 37°C. The reactions were stopped by the addition of Proteinase K (Roche) and incubated in a 60°C water bath for 25 min. Samples were then analyzed by agarose gel electrophoresis. Cleavage sites are distant from the target site for these enzymes; therefore, a computer program, RMsearch, was used to search for target sequences present in plasmids cut by the enzyme and not present in uncut plasmids (58).

## RESULTS

The four variants of the CC398-1 SauSTORF499P Type I RM system shown in Figure 1B and C were constructed. The nuclease fusion had ten subvariants with differing lengths of the 5' end of *hsdR* fused to the 5' end of *hsdM* as shown in Supplementary Table S2. These different lengths were chosen by comparison of the sequence of HsdR with the known sequence and structure of the HsdR protein from the EcoR124I RM system (59,60), Supplementary Figure S1. This process was informed by secondary sequence predictions and a protein model, created by the Phyre<sup>2</sup> online software (61).

Two subvariants of the M<sub>1/2</sub>S protein were engineered with the full length *hsdM* followed by the first 666bp or 639bp of *hsdS*. The longer sequence from *hsdS* encodes all of the central conserved region of HsdS. The M<sub>1/2</sub>S<sub>fus</sub> construct contained the first 666bp of *hsdS* fused in frame to *hsdM*.

### In vivo analysis of restriction and modification activity using phage λ

The *in vivo* assay detecting restriction and modification of λ phage is a simple way to assess the activity of the variants of the CC398-1 RM system. Efficiency of plating (eop) of phage λ is the ratio of phage titre on a restricting strain versus a non-restricting strain: in this case a strain transformed

with two plasmids, one for HsdR and one for the MTase, versus the same strain transformed with the MTase plasmid. The strain containing the 'wild-type' CC398-1 MTase plus the HsdR showed nearly a ten-fold reduction in eop and is therefore active in restriction, Table 1.

No significant change in eop from a value of one (values range from 0.5 to 2.25) was observed in strains expressing any of the ten nuclease-HsdM fusions along with HsdS, Table 1, indicating that none of these fusions possess restriction activity. However, phage recovered from these assays had all become modified as shown by the titre of these phage on the strain harbouring the wild-type CC398-1 MTase + HsdR compared to their titre on a strain lacking HsdR, Table 1. This indicates that these fusions are active in modification and have the same sequence specificity as the wild-type system. However as described in the next section, the protein instability noted during purification may also be present *in vivo* so the modification activity may be due to an assembly without the nuclease domain, a possibility that is difficult to rule out at present.

In the presence of HsdR, the  $M_{1/2}S$  protein was only active in modification but not restriction, Table 1. This is in contrast to the situations found with other truncations of the S subunit of Type I RM systems (28,29,31). This may be due to protein instability as observed when purifying this protein as described below. In contrast, both  $MS_{fus}$  and  $M_{1/2}S_{fus}$  in the presence of HsdR were active in both modification and restriction, Table 1. The apparent regaining of restriction activity when comparing  $M_{1/2}S$  with  $M_{1/2}S_{fus}$  is presumably due to an improved protein stability as discussed below as the amount of HsdR subunit being expressed in the cells should be the same.

### Protein expression

In our attempts to make a Type IIB RM enzyme, only one of the nuclease domain-HsdM fusions, RM\_EB.2 protein, could be expressed and partially purified along with the S subunit, Supplementary Figure S2. SDS-PAGE of fractions from a size exclusion chromatography column indicated that the HsdS initially coeluted with a small amount of RM\_EB.2 (lane 4 in Supplementary Figure S2) followed by coelution of HsdS with RM\_EB.2 and a ~70 kDa fragment (see lanes 6 to 8 in Supplementary Figure S2). This was then followed by coelution of HsdS with RM\_EB.2 and a fragment of ~61 kDa (see lanes 8 and 9 in Supplementary Figure S2) and lastly by coelution of HsdS with RM\_EB.2 and a fragment of ~65 kDa (see lane 10 in Supplementary Figure S2). As it is known that deletion of the C-terminus of the HsdM subunit prevents assembly with HsdS (62) then these fragments of RM\_EB.2 must be deletions of the N-terminal region and their size indicates that these fragments have lost the nuclease domain. Thus, it appears that RM\_EB.2 predominantly forms a complex of the form RM\_EB.2 + truncated fragment + HsdS and this may explain the absence of nuclease activity *in vivo* as the complex mostly contains only a single nuclease domain.

The  $MS_{fus}$ ,  $M_{1/2}S$  and  $M_{1/2}S_{fus}$  fusions and truncations all expressed well and were soluble, Supplementary Figures S3, S4 and S5, respectively. Although it had a tendency to aggre-

gate with time, the  $M_{1/2}S_{fus}$  protein was the purest of these three proteins.

The  $MS_{fus}$  protein was unstable during the purification procedures and underwent some proteolysis. The additional bands on an SDS-PAGE gel were excised and sent for analysis by peptide fragmentation mass spectrometry at the Mass Spectrometry and Proteomics Facility (University of St. Andrews). Mass spectrometry results confirmed the larger of the contaminating species contained a C-terminally clipped form of the  $MS_{fus}$  species of ~73 kDa and the smaller species was an unrelated *E. coli* protein (data not shown). The ~73 kDa fragment comprises all of HsdM and about half of the first TRD of HsdS. A Sephacryl S200 size exclusion purification step was added to the purification method, however the larger fragment remained bound to the  $MS_{fus}$  suggesting that a proportion of the CC398-1  $MS_{fus}$  protein is copurified with the clipped fragment.

Analysis of pooled protein fractions on a calibrated analytical size exclusion column showed a single major elution peak for CC398-1 MTase,  $M_{1/2}S$  and  $M_{1/2}S_{fus}$  proteins with apparent molecular masses of 241, 225 and 251 kDa respectively, Supplementary Figure S6 and Supplementary Table S5. The expected molecular masses for these proteins are 166 to 173 kDa assuming the forms  $M_2S_1$ ,  $M_2(1/2S)_2$  and  $M(1/2S_{fus})_2$  for these proteins. It has been previously noted that Type I MTases elute with this higher than expected molecular mass due to their highly non-spherical shape (22,63–69). The  $M_{1/2}S$  and  $M_{1/2}S_{fus}$  proteins also showed material eluting before the main protein peak indicating the presence of very high molecular mass aggregates. As noted above, the  $MS_{fus}$  protein co-eluted with a proteolytic fragment from the Sephacryl S200 column. The analytical column showed a complex asymmetric elution profile with the maximum UV absorption corresponding to a molecular mass of 325 kDa, Supplementary Figure S6. This higher mass may indicate that the fusion protein can also exist as a dimer ( $MS_{fus})_2$ . The broadness of the elution profile may also suggest monomeric fusion protein and the existence of a complex equilibrium of different quaternary structures.

### Plasmid cleavage activity in vitro

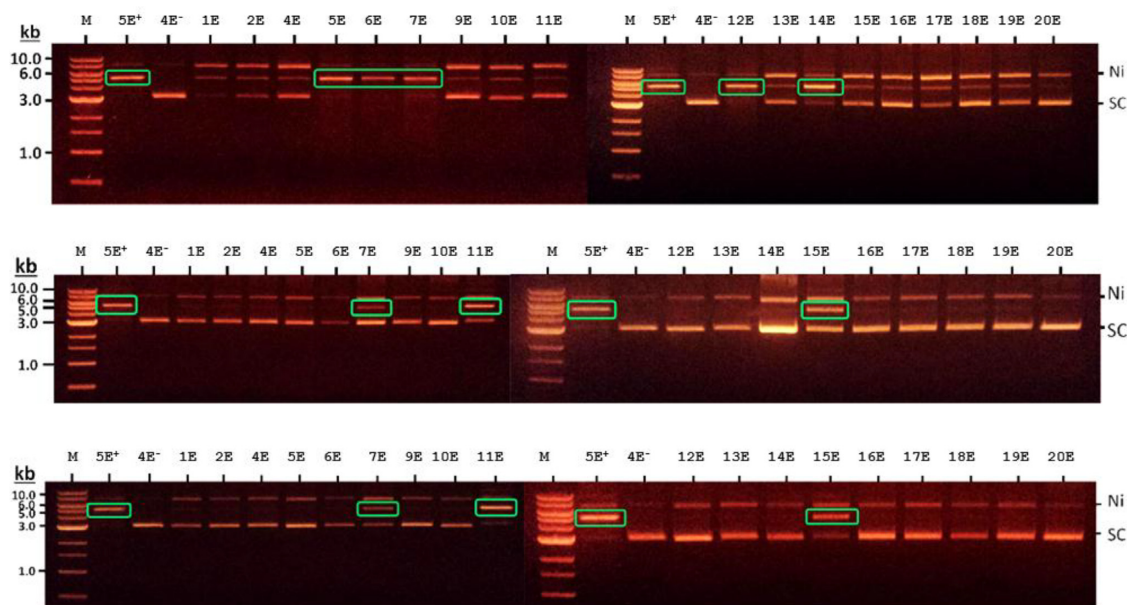
A set of plasmids containing inserts of known sequence was used for cleavage assays. Conversion of closed circle, supercoiled plasmid DNA to linear form was used as evidence of cleavage. As previously noted, the introduction of single-strand nicks into the plasmids to give an open circular form is attributed to the presence of nuclease contaminants in our enzyme preparations and is not evidence for the presence of a target site (27). In addition, the  $MS_{fus}$  enzyme produced a small amount of linearised DNA with every plasmid used, Figure 2 top row. However, in this case the absence or near absence of closed circular DNA in plasmids containing the target site was obvious and gave a target agreeing with the results of SMRT sequencing.

The partially purified complex of nuclease-HsdM fusion, RM\_EB.2, and S subunit showed no cleavage activity when incubated with a plasmid containing a target site (data not shown). This is in agreement with the *in vivo* results shown in Table 1.



**Table 1.** The *in vivo* restriction and modification activity of the wild type CC398-1 Type I RM system and its derivatives assessed by the efficiency of plating (eop) of phage  $\lambda_v$  on *E. coli* strain NM1261 transformed with plasmids expressing CC398-1 (HsdR + HsdM + HsdS) or its derivatives (first column). NM1261 has no RM activity overlapping with RM systems in this investigation. The first column also shows the length of the nucleotide sequences taken from the 5' end of the *hsdR* gene for fusion to the 5' end of the *hsdM* gene, and the equivalent number of amino acids (aa), to make the RM\_EB\_X series of fusions. Two versions of the  $M_{1/2}S$  construct were tested. The second column shows the eop of phage infecting the strain transformed with the RM system relative to the strain cotransformed with MTase and pRSFDuet-1 (the vector used for supplying *hsdR*). A low eop ( $\sim 0.1$ ) indicates restriction proficiency. The third column shows the eop of phage recovered from the experiments in the second column on reinfection of strains transformed with the RM system relative to the strain cotransformed with MTase and pRSFDuet-1. A high eop ( $\sim 1$ ) indicates modification proficiency

RM system	Eop of phage $\lambda_v$ on <i>E. coli</i> NM1261 expressing the system	Eop of recovered phage $\lambda_v$ on <i>E. coli</i> NM1261 expressing the system
HsdR + HsdM + HsdS	0.15	0.88
HsdS + RM_EB.2 (1–786, aa 1–262)	1.1	1.67
HsdS + RM_EB.3 (1–651, aa 1–217)	0.68	0.8
HsdS + RM_EB.4 (1–621, aa 1–207)	1.5	1.13
HsdS + RM_EB.5 (1–645, aa 1–215)	2.25	1.06
HsdS + RM_EB.6 (1–675, aa 1–225)	1.0	0.67
HsdS + RM_EB.7 (1–732, aa 1–244)	2.0	1.0
HsdS + RM_EB.8 (1–513, aa 1–171)	0.75	0.89
HsdS + RM_EB.9 (1–549, aa 1–183)	1.25	0.8
HsdS + RM_EB.10 (1–816, aa 1–272)	0.75	0.59
HsdS + RM_EB.11 (1–855, aa 1–285)	0.5	1.25
HsdR + $MS_{fus}$	0.09	0.78
HsdR + $M_{1/2}S$ ( $M_{1/2}S$ shorter version)	0.58 (0.40)	0.75 (0.70)
HsdR + $M_{1/2}S_{fus}$	0.10	1.02



**Figure 2.** Agarose gel electrophoresis analysis of samples from a plasmid cleavage assay using the 'E' series of plasmids described previously (27). From top to bottom the gels show the effect of the purified HsdR incubated with  $MS_{fus}$ ,  $M_{1/2}S$  and  $M_{1/2}S_{fus}$  on the plasmid. The gels show three different plasmid DNA species. The species that occurs at the top of each gel lane has a single strand cut and is 'nicked' (Ni). Below this is the linearised species, when both strands have been cut by enzymatic activity (highlighted by green boxes). The lowest species is the uncleaved supercoiled DNA (SC). The lanes marked M are molecular size markers (kb). Lanes marked 5E<sup>+</sup> and 4E<sup>-</sup> are plasmids incubated with HsdR plus CC398-1 MTase that contain a single site or no site for the RE and act as positive and negative controls, respectively.

The  $MS_{fus}$ ,  $M_{1/2}S$  and  $M_{1/2}S_{fus}$  proteins were all active in cleavage when complemented with HsdR as shown in Figure 2. Of note is that extra HsdM was NOT required for  $MS_{fus}$  to be active. This is in contrast to the  $MS_{fus}$  protein constructed from the EcoKI Type I RM system which needed additional HsdM to be added to the reaction (42). This is consistent with the identification of a fragment of the fusion protein remaining associated with CC398-1  $MS_{fus}$  throughout protein purification. RMsearch analy-

sis showed recognition by  $MS_{fus}$  of the wild-type CC398-1 RM enzyme target of ACCN<sub>6</sub>TGA as anticipated (26). The  $M_{1/2}S$  proteins,  $M_{1/2}S$  and  $M_{1/2}S_{fus}$  recognise a different target, ACCN<sub>5</sub>GGT, which is palindromic and has a spacer change resulting in the adenine targets being one base pair further apart than for the wild type CC398-1 RM enzyme (9 bp rather than 8 bp) even though the overall target sequence is one base pair shorter (11 bp rather than 12 bp). The DNA cleavage activity of  $M_{1/2}S_{fus}$  and  $M_{1/2}S$  contrasts



with the *in vivo* restriction activity, which was only clearly observed for  $M_{1/2}S_{fus}$ , suggesting an impaired association of  $M_{1/2}S$  with HsdR or too low an amount of HsdR being expressed *in vivo*.

### Single-molecule real-time (SMRT) sequencing

SMRT sequencing is able to identify modified bases in sequenced DNA and can therefore be used to identify the DNA recognition motif of a Type I MTase (e.g. (25,26)). The method is particularly useful in that all modification is conducted *in vivo* eliminating the need for large amounts of purified protein.

The expression plasmids containing the wild-type CC398-1 MTase,  $MS_{fus}$ ,  $M_{1/2}S$  and  $M_{1/2}S_{fus}$  genes were used to transform competent *Escherichia coli* ER2796 cells, a strain with no inherent DNA methylation. Genomic DNA was isolated from individual *E. coli* colonies expressing the protein of interest. The results from SMRT sequencing, Table 2 and Supplementary Figure S7, confirmed that the manipulated MTase genes were expressing active MTases *in vivo* in agreement with the results in Table 1 and Figure 2. The DNA recognition sequence for  $MS_{fus}$  was  $ACCN_5RTGA$  as expected although some methylation of  $ACCN_5RTGG$  was also detected. Analysis of  $M_{1/2}S$  and  $M_{1/2}S_{fus}$  show the novel enzymes methylating the palindromic sequence,  $ACCN_5GGT$ . This finding also confirmed that the  $M_{1/2}S$  and  $M_{1/2}S_{fus}$  dimerise to recognise the palindromic target sequence.

### DISCUSSION

By making step-wise alterations to the structure of the CC398-1 Type I RM system, novel, soluble, sequence-specific enzymes have been engineered. As previously assumed from sequence comparisons (8), our results indicate that the Type I RM enzymes are close relatives of the IIB and IIG RM enzymes in terms of the organisation of their structural domains.

Although not restriction active, fusions of the nuclease domain of HsdR to the HsdM could be produced and purified in an assembly with the S subunit. Most of the purified material is a complex of RM\_EB\_2 with HsdS and a fragment of RM\_EB\_2 lacking the nuclease domain. This complex retains methylation activity but lacks nuclease activity suggesting that two nuclease domains are necessary. However, the successful expression and assembly of the fusion protein with HsdS suggests that our approach eventually could create a viable RM system, but that different lengths of HsdR would need to be fused to HsdM and tested before a fully active analogue of a Type IIB RM enzyme is obtained. To achieve full restriction and modification activity will require more optimisation but as the proteins are expressed, soluble and bind to the S subunit this is technically feasible although probably very challenging.

Manipulation of the *hsdM* and *hsdS* genes proved more successful and represents a step towards a Type IIG RM system. Three variants with HsdM fused to HsdS ( $MS_{fus}$ ), a half HsdS which dimerised to produce a 'complete' HsdS subunit ( $M_{1/2}S$ ) and a fusion of the half HsdS to HsdM which also dimerised ( $M_{1/2}S_{fus}$ ) were all active MTases and,

when complemented with HsdR, active restriction endonucleases. The modification was efficient *in vivo* as shown by the eop assay and the high motif coverage observed in the SMRT sequencing results. The target sites were as predicted from knowledge of the TRD specificity although the number of base pairs separating the methylation sites was increased by one base pair. The restriction *in vivo* was poor given the number of targets on phage  $\lambda$  (12 for  $ACCN_5NNRTGA$  and 14 for  $ACCN_5NNNNNGGT$ ) but in line with previous results with the Type I RM systems from *S. aureus* being expressed in *E. coli* (25–27). The observed eop for the restriction positive strains is poor when compared to the much higher levels of cutback observed with, for example, the EcoKI Type I RM system ( $eop = 10^{-5}$  typically) (42). This poor eop may be due to poor protein stability, inadequate expression of the HsdR subunits compared to the MTase subunits or a lack of target sites on the phage. Given the number of targets on phage  $\lambda$  for these enzymes it is more probable that poor stability, especially for  $M_{1/2}S$ , and a mismatch in the intracellular concentrations of the HsdR and MTase proteins is responsible for the poor eop.

An unexpected result was observed in the new palindromic sequence, recognised by the dimerising  $M_{1/2}S$  and  $M_{1/2}S_{fus}$  fusion proteins. The sequence is a nucleotide shorter than that recognised by the wild-type enzyme, but the methylated bases are one bp further apart. This effect was not seen in the half HsdS enzymes created by others (28,29). The main role of the central conserved region has been assumed to be in separating the TRDs to a specific distance to recognise the bipartite target sequence but the new target spacing suggests that there is not a simple relationship between the length of the central conserved region and the DNA target.

In theory, one final step could be taken to fuse the nuclease domain of HsdR to the  $M_{1/2}S_{fus}$  fusion protein to give an arrangement of domains identical to a Type IIG RM enzyme shown in Figure 1A. Our results would suggest that such a protein would be expressed and active as a MTase but that the nature of the fusion of the nuclease domain to MTase would need much optimisation in the laboratory to produce a fully functional RM enzyme. Over long evolutionary time periods, Nature has presumably made many attempts to make the Type IIG RM enzymes by fusing the nuclease domain of HsdR to the  $M_{1/2}S_{fus}$  fusion protein. Most of these fusions have failed to function and been eliminated but at least one attempt must have succeeded and spread through bacterial populations such that today Type IIG RM systems are almost as prevalent as the Type I RM systems, Table 3.

### A structural model for the evolution of RM systems

This work aimed to alter a Type I RM system to structural forms resembling 'simpler' Types of RM systems to provide evidence for the theory that the RM Types are evolutionarily linked. The Type I, IIG and III systems are by far the most common RM systems in bacteria and archaea and are well conserved in terms of genetic structure and protein structure reusing the same domains in different combinations (5,8,16,70), Figure 1A. Table 3 shows the relative proportions of the different RM Types with Type I, IIG and

**Table 2.** SMRT results determining the sites of methylation on the genomic DNA of *E. coli* ER2796 using cells transformed with plasmids expressing the wild type CC398 MTase, MS<sub>fus</sub>, M<sub>1/2</sub>S or M<sub>1/2</sub>S<sub>fus</sub> proteins. MS<sub>fus</sub> recognises the same target as the CC398 MTase as expected but also modifies some sites with the second part of the site having the sequence HRTGG rather than RTGA

Enzyme	Motifs	m6A modified position	% Motifs detected	Number of motifs detected	Number of motifs on genome
CC398 MTase	ACNNNNNNRTGA	1	99.69	971	974
	TCAYNNNNNGGT	3	99.69	971	974
MS <sub>fus</sub>	ACNNNNNNRTGA	1	76.88	1663	2163
	ACNNNNNHRTGG	1	69.46	523	753
	TCAYNNNNNGGT	3	76.05	1645	2163
M <sub>1/2</sub> S	ACNNNNNNGGT	1	78.64	2095	2664
M <sub>1/2</sub> S <sub>fus</sub>	ACNNNNNNGGT	1	99.66	2655	2664

**Table 3.** The number of restriction enzymes, both confirmed and putative, in REBASE as at April 2018. The numbers were derived by simply counting the number of R subunits for the Type I RM systems, the number of REases for the Type II systems and Type IV systems and the number of res subunits for the Type III RM systems. The proportions are very similar to those obtained previously by Oliveira *et al.*, (5)

RM system	Number of examples	Percentage of total	Percentage of total ignoring Type IV
Type I REase	18580	26.6	31.8
Type IIP REase	13648	19.6	23.4
Type II REase (not IIP or IIG)	7144	10.2	12.2
Type IIG REase/MTase	11737	16.8	20.1
Type III REase	7305	10.5	12.5
Type IV REase	11356	16.3	-
Total	69770	100	100

III making up ~55% of known and putative RM systems in REBASE (6). The methylation-dependent Type IV restriction systems make up a further 16% leaving the most well-known Type IIP plus Type IIS and other unassigned Type II RM systems to make up the remaining ~30% of RM systems.

Our results linking different Types of RM system suggest an evolutionary model for the diversity of RM enzymes. The model, shown in Figure 3, is based upon suggesting answers to the questions posed in the introduction and makes specific proposals for experimental tests as described below.

When did RM first appear in evolutionary history and why did it appear?

It would seem very probable that RM systems appeared as soon as horizontal gene transfer (HGT) between early cells arose and they came in contact with the first phage or other mobile genetic elements (71–74). RM would help any host cell to control the rate of uptake of foreign DNA and this would probably be crucial to generate stability in the primitive genome. If HGT was not controlled, then it is difficult to see how a stable genome could be established and maintained in an environment with a lot of mobile genetic elements (MGE). Conversely too much control of HGT would slow down the subsequent evolution of the RM host as no foreign DNA on MGE could enter the host which would then have to rely solely upon vertical evolution.

What did the first RM system look like?

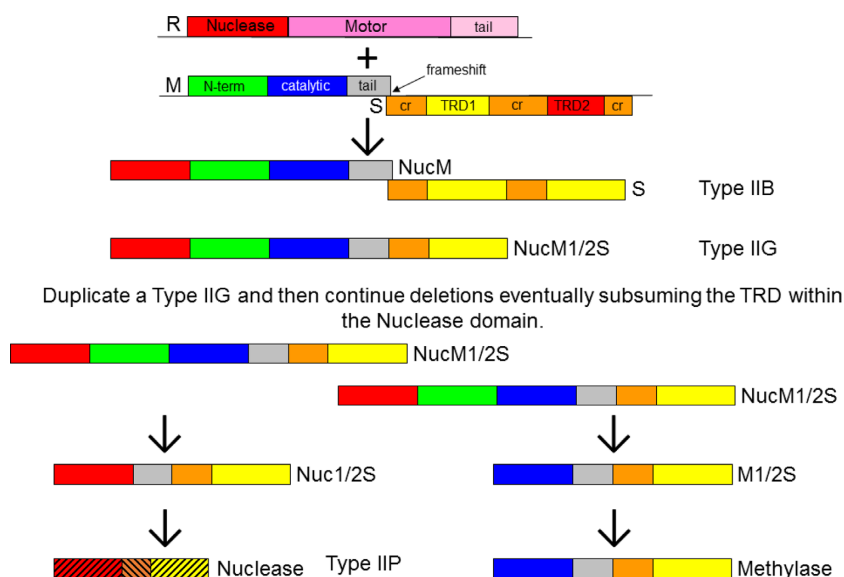
The first RM system would have to be constructed from whatever components were present in the early cell. It seems to us to be highly improbable that classical Type IIP RM systems with separate REase and MTase enzymes using entirely different chemistries to perform their separate reactions while recognising the exact same DNA target would evolve easily. The classical Type IIP RM systems are not

‘simple’ systems despite their small size in molecular weight terms.

We suggest that the first RM enzyme would be akin to a Type I RM enzyme as the subunits for these enzymes have close homologues in ancient essential enzymes and proteins vital for genome repair, nucleotide modification and regulation of transcription. When an early cell evolved a primitive RM defense system, it would have to build upon the components already present. Maintenance of an early genome would require DNA repair enzymes and these frequently contain ATP-dependent DEAD-box SFII superfamily proteins of which HsdR is a member (75). Nucleotide methylation utilises an ancient SAM-dependent fold and this fold is found in the MTases used in all RM systems (76,77). Gene regulation using transcription factors to recognise long palindromic sequences often use homodimeric proteins perhaps akin to the dimers of half-HsdS. For instance, the BmrR repressor recognises split DNA targets similar to the targets recognised by the Type I RM systems containing half-HsdS and its structure is strikingly similar to HsdS (78). When under attack from phage, any early cell that could assemble an RM system from these components would have an advantage.

Why did it subsequently evolve to form the large range of RM structural variants observed today?

This early ‘Type I’ RM enzyme built from pre-existing components would subsequently evolve to form other ‘simpler’ RM systems by the sorts of domain fusions/deletions investigated in this work and elsewhere (28,29) under the evolutionary pressure of antirestriction (79–81). Most antirestriction / antimodification (anti-RM) systems are known to be directed at the Type I RM systems (79). These anti-RM systems, such as DNA mimics (81), would provide the evolutionary pressure to force the early ‘Type I’ RM system to evolve to an anti-RM-resistant form. These



**Figure 3.** Schematic diagram delineating the steps that could be taken to alter the structure of a Type I RM system to create Type II systems. The colour scheme for domains is the same as in Figure 1. In the step from Nuc<sup>1/2</sup>S to Nuclease, the domains are merging together so are shown with hatching.

anti-RM-resistant forms would, by definition, be the Type II and Type III RM systems. It appears the Type II and Type III RM systems are not targeted by the existing DNA mimics but this clearly needs further investigation as so few have been tested for inhibition by a DNA mimic (82). The first step in this ‘simplification’ of the structure of the RM enzyme would appear to be via the evolution of the Type IIB and IIG RM enzymes by the gene fusions and deletions investigated here. Once a cell had evolved a RM system resistant to the MGE-borne DNA mimics, the MGE would then be under further pressure to evolve new anti-RM mechanisms and it appears that target site avoidance and nucleotide modifications are particularly prevalent choices (79,83).

What was the evolutionary pathway for the appearance of these structural variants?

If the proposed scenario is at least approximately correct then what happened after the appearance of the Type IIB/G enzymes to generate the Type IIP and Type III RM systems?

Type III RM systems, composed of ‘res’ and ‘mod’ subunits (9), would appear to use a variant of the HsdR subunit plus a fused HsdM-HsdS where there has been some circular permutation to move the *hsdS* to within the *hsdM* gene (17,18,43,53). The relative locations of the TRD with respect to the motifs for SAM binding subdivide all SAM-dependent MTases into six groups,  $\alpha$  to  $\zeta$ , with the  $\gamma$  group predominating (17,43). This circular permutation, with different relative locations of the amino acid motifs of the SAM-binding domain with respect to the TRD, has also occurred in some Type II RM systems with separate MTases leading to the formation of  $\alpha$  and  $\beta$  subgroups from the  $\gamma$  subgroup, which is the main subgroup in the Type I and IIB/G RM systems (17,43,84).

To generate separate REase and MTase as found in ‘classical’ Type IIP RM systems would require a gene duplication of the nascent Type IIG gene followed by evolution of the REase by deletion of the M domain and evolution of

the MTase by loss of the nuclease domain and in both cases retaining the TRD(s), Figure 3. At this point, due to the selfish nature of RM systems with separate R and M (11), the MTase would be under pressure to maintain sequence specificity while the REase could start to mutate to obscure the amino acid conservation present immediately after the duplication event. The REase could even acquire new functions (3).

Of course, one may ask why the Type IIP REase has two ‘TRD’ regions to recognise a palindromic target while the cognate MTase has only a single different TRD? This would seem to be a problem with our model. However, very few RM systems have structures known for both REase and MTase (PvuII seems to be the only pair apart from those with combined R and M) (85,86) and there is a regrettably little information on the quaternary structure of the active enzymes. Of note here is that many simple Type II MTases appear to exist as dimers despite the fact that they are usually assumed to operate as monomers (44). A dimeric MTase may have two TRDs but only use one at a time as this would allow them to simultaneously scan both DNA strands or to move around on DNA via DNA looping (87). Furthermore, many REases appear to require rather complicated quaternary structures to assemble on DNA to be active (88–92). Recent structures of some REases show distinct similarity to an HsdS subunit (54) and this would support our suggestion that a proto-IIG enzyme lost its MTase domain resulting in a fusion of a nuclease domain to a half-HsdS domain which would then dimerise to produce a IIP REase with a homodimeric structure and a palindromic target (the cognate MTase would lose the nuclease domain instead). One of these REases, R.SwaI, recognises 5'-ATTAAAT-3' and the two long alpha helices forming the dimer interface have their helical axes almost perpendicular to the DNA helical axis (54), Supplementary Figure S8. This arrangement places the two nuclease domains close to the centre of the target to allow cutting at the centre



to produce blunt ends. A similar arrangement is found with R.HincII which recognises GTYRAC and makes a blunt cut (91,92), Supplementary Figure S8. In comparison, models of HsdS bound to DNA (22,24) have the equivalent alpha helices aligned nearly parallel with the DNA helical axis and this places the TRDs far apart to allow recognition of the classic bipartite Type I target sequence, Supplementary Figure S8. Perhaps by changing the angle between helical axes of the alpha helices and the DNA helix, Nature has been able to bring the DNA recognition domains of a proto-half-HsdS closer together to recognise shorter targets and to incorporate the necessary features of the nuclease domain to facilitate DNA cleavage? Even Type IIP REases without such an obvious structural relationship to an HsdS use extensive alpha helical regions to form the interface between the two subunits in a similar manner to the interface formed between two half-HsdS monomers (12).

Our model may also suggest why C5C methylation uses a very highly conserved MTase domain and why it is found predominantly in Type II MTases and hardly ever in other RM Types. Type I, IIB and IIG MTases, almost without exception, are N6A or N4C MTases (93). Some are capable of methylating both as they can be tricked into methylating the wrong nucleotide (85) but they do share a common MTase domain architecture which is very ancient (76). The MTase N6A/N4C catalytic domain, although having relatively few conserved amino acid motifs, has a well conserved fold and catalytic site (7,43,70). The MTase domain in a C5C MTase has exactly the same fold as the N6A and N4C MTases but the amino acid sequence is much better conserved and 10 motifs, including motifs equivalent to those in N6A and N4C MTases, are easily recognised (7,43,70). This suggests that C5C methylation is a more recent innovation in evolution and that it is a difficult chemistry to evolve as it would require multiple, simultaneous amino acid changes in a precursor N4C MTase domain (12). Thus, C5C methylation chemistry probably appeared once and, by chance, became the main methylation found in eukaryotes.

### Further experiments suggested by the structural model

Previously evolutionary models have been proposed for the Type II RM systems (17,95,96) and similarities between several subgroups of the Type II RM systems and the Type I and III RM systems have been noted (7,8,45–49). However, when our results are coupled with these evolutionary models it is possible to suggest a ‘global’ model the majority of the defined ‘Types’ of RM system (excepting the Type IV restriction systems). Of course, our model shown in Figure 3 is not all-encompassing as there are other Type II RM variants such as IIS and those using other nuclease chemistries which exist, but these are a small proportion of the total number of RM systems (5,6). However, our model covers the great majority of known RM systems (5).

The model, even if subsequently proven to be incorrect, highlights several areas of RM research that have been neglected and which should be investigated. These areas are listed below. All of these suggestions are readily addressed due to easy availability of purified commercial RM enzymes.

Perhaps surprisingly, there are very few structures of REases and MTases from the same RM system (85,86). More RM ‘pairs’ from Type IIP RM systems should be solved to see if they use similar structural features to recognise their DNA target sequence. In other words, do the separate REase and MTase show any sign of a conserved ancestral TRD?

Anti-RM DNA mimics have never been tested against the majority of RM systems other than the Type I RM systems (82). Given the commercial availability of many REases and MTases, anti-RM versus the Type II RM enzymes could also easily be investigated to see if the DNA mimics targeting the Type I RM enzymes are more versatile than currently assumed. The phage T7 ocr DNA mimic (97,98) and the ArdA DNA mimic from conjugative plasmids and transposons (56,57,99) could be easily employed for this purpose but further anti-RM proteins such as T3 SAMase (100), Ral (101) and Lar (102) could be investigated.

The quaternary structure of the RM enzymes when bound to DNA needs to be examined as, in the few examples studied, drastic changes in their structure upon binding have been observed (91). Electron microscopy would seem to be a particularly useful technique in this area (91), especially if applied to the large protein-DNA complexes required for the REase activity of the Type IIB RM enzymes (89). Atomic force microscopy, although of lower resolution than electron microscopy, may also be useful for elucidating DNA topology in these complexes (90,103–105).

The evolution of methylation of cytosine at the C5 position is clearly difficult as it appears to have only occurred relatively recently and only in the Type II RM systems. Assuming that the predecessor of C5C methylation was N4C methylation, it may be possible to mutate the catalytic site in a N4C MTase to recognise and methylate C5C (or *vice versa* which may be an easier experiment to pursue in practice) with a few amino acid changes. In this vein, and as suggested by the structure of the SmaI REase (54), it may also be possible to mutate a TRD in a MTase or HsdS subunit to incorporate the PD-(D/E)XK nuclease motif and generate a new REase.

### SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

### ACKNOWLEDGEMENTS

D.T.F.D. thanks the Institute of Advanced Study, Durham University for providing a fellowship from January to April 2016 and an excellent environment for initiating this paper. We gratefully acknowledge assistance from Dana Macelis (New England Biolabs) in compiling the information necessary for constructing Table 3.

### FUNDING

Biotechnology and Biological Sciences Research Council (BBSRC) [BB/K005804/1 to D.T.F.D.]. Funding for open access charge: BBSRC

*Conflict of interest statement.* None declared.



## REFERENCES

- Thomas, C.M. and Nielsen, K.M. (2005) Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat. Rev. Microbiol.*, **3**, 711–721.
- Makarova, K.S., Wolf, Y.I. and Koonin, E.V. (2013) Comparative genomics of defense systems in archaea and bacteria. *Nucleic Acids Res.*, **41**, 4360–4377.
- Vasu, K. and Nagaraja, V. (2013) Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiol. Mol. Biol. Rev.*, **77**, 53–72.
- van Houte, S., Buckling, A. and Westra, E.R. (2016) Evolutionary ecology of prokaryotic immune mechanisms. *Microbiol. Mol. Biol. Rev.*, **80**, 745–763.
- Oliveira, P.H., Touchon, M. and Rocha, E.P. (2014) The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Res.*, **42**, 10618–10631.
- Roberts, R.J., Vincze, T., Posfai, J. and Macelis, D. (2015) REBASE - a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.*, **43**, D298–D299.
- Loenen, W.A.M., Dryden, D.T.F., Raleigh, E.A., Wilson, G.G. and Murray, N.E. (2014) Highlights of the DNA cutters: a short history of the restriction enzymes. *Nucleic Acids Res.*, **42**, 3–19.
- Loenen, W.A.M., Dryden, D.T.F., Raleigh, E.A. and Wilson, G.G. (2014) Type I restriction enzymes and their relatives. *Nucleic Acids Res.*, **42**, 20–44.
- Rao, D.N., Dryden, D.T.F. and Bheemanaik, S. (2014) Type III restriction-modification enzymes: a historical perspective. *Nucleic Acids Res.*, **42**, 45–55.
- Loenen, W.A.M. and Raleigh, E.A. (2014) The other face of restriction: modification-dependent enzymes. *Nucleic Acids Res.*, **42**, 56–69.
- Mruk, I. and Kobayashi, I. (2014) To be or not to be: regulation of restriction-modification systems and other toxin-antitoxin systems. *Nucleic Acids Res.*, **42**, 70–86.
- Pingoud, A., Wilson, G.G. and Wende, W. (2014) Type II restriction endonucleases—a historical perspective and more. *Nucleic Acids Res.*, **42**, 7489–7527.
- Levin, B.R. (1988) Frequency-dependent selection in bacterial populations. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, **319**, 459–472.
- Korona, R. and Levin, B.R. (1993) Phage-mediated selection and the evolution and maintenance of Restriction-Modification. *Evolution*, **47**, 556–575.
- Lindsay, J.A. (2014) *Staphylococcus aureus* genomics and the impact of horizontal gene transfer. *Intl. J. Med. Microbiol.*, **304**, 103–109.
- Oliveira, P.H., Touchon, M. and Rocha, E.P. (2016) Regulation of genetic flux between bacteria by restriction-modification systems. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 5658–5663.
- Bujnicki, J. M. (2001) Understanding the evolution of restriction-modification systems: Clues from sequence and structure comparisons. *Acta Biochim. Pol.*, **48**, 935–967.
- Roberts, R.J., Belfort, M., Bestor, T., Bhagwat, A.S., Bickle, T.A., Bitinaite, J., Blumenthal, R.M., Degtyarev, S.K., Dryden, D.T.F., Dybvig, K. *et al.* (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res.*, **31**, 1805–1812.
- Roberts, R. J. (2005) How restriction enzymes became the workhorses of molecular biology. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 5905–5908.
- Murray, N. E. (2000) Type I restriction systems: sophisticated molecular machines (a Legacy of Bertani and Weigle). *Microbiol. Mol. Biol. Rev.*, **64**, 412–434.
- Šišáková, E., Weiserová, M., Dekker, C., Seidel, R. and Szczelkun, M. D. (2008) The interrelationship of helicase and nuclease domains during DNA translocation by the molecular motor EcoR124I. *J. Mol. Biol.*, **384**, 1273–1286.
- Kennaway, C.K., Obarska-Kosinska, A., White, J.H., Tuszyńska, I., Cooper, L.P., Bujnicki, J.M., Trinick, J. and Dryden, D.T.F. (2009) The structure of M.EcoKI Type I DNA methyltransferase with a DNA mimic antirestriction protein. *Nucleic Acids Res.*, **37**, 762–770.
- Roberts, G. A., Cooper, L. P., White, J. H., Su, T.-J., Zipprich, J. T., Geary, P., Kennedy, C. and Dryden, D. T. F. (2011) An investigation of the structural requirements for ATP hydrolysis and DNA cleavage by the EcoKI Type I DNA restriction and modification enzyme. *Nucleic Acids Res.*, **39**, 7667–7676.
- Kennaway, C.K., Taylor, J.E., Song, C.F., Potrzebowski, W., Nicholson, W., White, J.H., Swiderska, A., Obarska-Kosinska, A., Callow, P., Cooper, L.P. *et al.* (2012) Structure and operation of the DNA-translocating Type I DNA restriction enzymes. *Genes Dev.*, **26**, 92–104.
- Cooper, L.P., Roberts, G.A., White, J.H., Luyten, Y.A., Bower, E.K.M., Morgan, R.D., Roberts, R.J., Lindsay, J.A. and Dryden, D.T.F. (2017) DNA target recognition domains in the Type I restriction and modification systems of *Staphylococcus aureus*. *Nucleic Acids Res.*, **45**, 3395–3406.
- Chen, K., Stephanou, A.S., Roberts, G.A., White, J.H., Cooper, L.P., Houston, P.J., Lindsay, J.A. and Dryden, D.T.F. (2016) The Type I restriction enzymes as barriers to horizontal gene transfer: determination of the DNA target sequences recognised by livestock-associated methicillin-resistant *Staphylococcus aureus* Clonal Complexes 133/ST771 and 398. *Adv. Exp. Med. Biol.*, **915**, 81–97.
- Roberts, G.A., Houston, P.J., White, J.H., Chen, K., Stephanou, A.S., Cooper, L.P., Dryden, D.T.F. and Lindsay, J.A. (2013) Impact of target site distribution for Type I restriction enzymes on the evolution of methicillin-resistant *Staphylococcus aureus* (MRSA) populations. *Nucleic Acids Res.*, **41**, 7472–7484.
- Abadjieva, A., Patel, J., Webb, M., Zinkevich, V. and Firman, K. (1993) A deletion mutant of the type IC restriction endonuclease EcoR124I expressing a novel DNA specificity. *Nucleic Acids Res.*, **21**, 4435–4443.
- Meister, J., MacWilliams, M., Hübner, P., Jütte, H., Skrzypek, E., Piekarczyk, A. and Bickle, T.A. (1993) Macroeolution by transposition: drastic modification of DNA recognition by a type I restriction enzyme following Tn5 transposition. *EMBO J.*, **12**, 4585–4591.
- Kneale, G.G. (1994) A Symmetrical Model for the Domain Structure of Type I DNA Methyltransferases. *J. Mol. Biol.*, **243**, 1–5.
- MacWilliams, M.P. and Bickle, T.A. (1996) Generation of new DNA binding specificity by truncation of the type IC EcoDXXI hsdS gene. *EMBO J.*, **15**, 4775–4783.
- Janscak, P. and Bickle, T.A. (1998) The DNA recognition subunit of the type IB restriction-modification enzyme EcoAI tolerates circular permutations of its polypeptide chain. *J. Mol. Biol.*, **284**, 937–948.
- Calisto, B.M., Pich, O.Q., Piñol, J., Fita, I., Querol, E. and Carpena, X. (2005) Crystal structure of a putative type I restriction-modification S subunit from *Mycoplasma genitalium*. *J. Mol. Biol.*, **351**, 749–762.
- Kim, J.S., De Giovanni, A., Jancarik, J., Adams, P.D., Yokota, H., Kim, R. and Kim, S.H. (2005) Crystal structure of DNA sequence specificity subunit of a type I restriction-modification enzyme and its functional implications. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 3248–3253.
- Gao, P., Tang, Q., An, X., Yan, X. and Liang, D. (2011) Structure of HsdS subunit from *Thermoanaerobacter tengcongensis* sheds lights on mechanism of dynamic opening and closing of type I methyltransferase. *PLoS One*, **6**, e17346.
- Furuta, Y., Kawai, M., Uchiyama, I. and Kobayashi, I. (2011) Domain movement within a gene: a novel evolutionary mechanism for protein diversification. *PLoS One*, **6**, e18819.
- Price, C., Lingner, J., Bickle, T.A., Firman, K. and Glover, S.W. (1989) Basis for changes in DNA recognition by the EcoR124 and EcoR124/3 type I DNA restriction and modification enzymes. *J. Mol. Biol.*, **205**, 115–125.
- Gubler, M., Bragaglia, D., Meyer, J., Piekarczyk, A. and Bickle, T.A. (1992) Recombination of constant and variable modules alters DNA sequence recognition by type IC restriction-modification enzymes. *EMBO J.*, **11**, 233–240.
- Adamczyk-Popławska, M., Kondrzycka, A., Urbanek, K. and Piekarczyk, A. (2003) Tetra-amino-acid tandem repeats are involved in HsdS complementation in type IC restriction-modification systems. *Microbiology*, **149**, 3311–3319.
- Adamczyk-Popławska, M., Lower, M. and Piekarczyk, A. (2011) Deletion of one nucleotide within the homonucleotide tract present in the hsdS gene alters the DNA sequence specificity of type I restriction-modification system NgoAV. *J. Bacteriol.*, **193**, 6750–6759.

41. Sain, B. and Murray, N.E. (1980) The hsd (host specificity) genes of *E. coli* K 12. *Mol. Gen. Genet.*, **180**, 35–46.
42. Roberts, G.A., Chen, K., Cooper, L.P., White, J.H., Blakely, G.W. and Dryden, D.T.F. (2012) Removal of a frameshift between the hsdM and hsdS genes of the EcoKI Type IA DNA restriction and modification system produces a new type of system and links the different families of Type I systems. *Nucleic Acids Res.*, **40**, 10916–10924.
43. Wilson, G.G. and Murray, N.E. (1991) Restriction and Modification Systems. *Annu. Rev. Genet.*, **25**, 585–627.
44. Malygin, E.G., Evdokimov, A.A. and Hattman, S. (2009) Dimeric/oligomeric DNA methyltransferases: an unfinished story. *Biol. Chem.*, **390**, 835–844.
45. Jurenaite-Urbanaviciene, S., Serksnaite, J., Kriukiene, E., Giedriene, J., Venclovas, C. and Lubys, A. (2007) Generation of DNA cleavage specificities of type II restriction endonucleases by reassortment of target recognition domains. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 10358–10363.
46. Marshall, J.J.T. and Halford, S.E. (2010) The Type IIB restriction endonucleases. *Biochem. Soc. Trans.*, **38**, 410–416.
47. Nakonieczna, J., Kaczorowski, T., Obarska-Kosinska, A. and Bujnicki, J.M. (2009) Functional analysis of MmI from methanol utilizer *Methylophilus methylotrophus*, a subtype IIC restriction-modification enzyme related to type I enzymes. *Appl. Environ. Microbiol.*, **75**, 212–223.
48. Shen, B.W., Xu, D., Chan, S.H., Zheng, Y., Zhu, Z., Xu, S.Y. and Stoddard, B.L. (2011) Characterization and crystal structure of the type IIG restriction endonuclease RM.BpuSI. *Nucleic Acids Res.*, **39**, 8223–8236.
49. Callahan, S.J., Luyten, Y.A., Gupta, Y.K., Wilson, G.G., Roberts, R.J., Morgan, R.D. and Aggarwal, A.K. (2016) Structure of type III Restriction-Modification enzyme MmI in complex with DNA has implications for engineering new specificities. *PLoS Biol.*, **14**, e1002442.
50. Madsen, A. and Josephsen, J. (2001) The LlaGI restriction and modification system of *Lactococcus lactis* W10 consists of only one single polypeptide. *FEMS Microbiol. Lett.*, **200**, 91–96.
51. Chand, M.K., Nirwan, N., Diffin, F.M., van Aelst, K., Kulkarni, M., Pernstich, C., Szczelkun, M.D. and Saikrishnan, K. (2015) Translocation-coupled DNA cleavage by the Type ISP restriction-modification enzymes. *Nat. Chem. Biol.*, **11**, 870–877.
52. Kulkarni, M., Nirwan, N., van Aelst, K., Szczelkun, M.D. and Saikrishnan, K. (2016) Structural insights into DNA sequence recognition by Type ISP restriction-modification enzymes. *Nucleic Acids Res.*, **44**, 4396–4408.
53. Gupta, Y.K., Chan, S.H., Xu, S.Y. and Aggarwal, A.K. (2015) Structural basis of asymmetric DNA methylation and ATP-triggered long-range diffusion by EcoP15I. *Nat. Commun.*, **6**, 7363.
54. Shen, B.W., Heiter, D.F., Lunnen, K.D., Wilson, G.G. and Stoddard, B.L. (2017) DNA recognition by the SwI restriction endonuclease involves unusual distortion of an 8 base pair A:T-rich target. *Nucleic Acids Res.*, **45**, 1516–1528.
55. Anton, B.P., Mongodin, E.F., Agrawal, S., Fomenkov, A., Byrd, D.R., Roberts, R.J. and Raleigh, E.A. (2015) Complete Genome Sequence of ER2796, a DNA Methyltransferase-Deficient Strain of *Escherichia coli* K-12. *PLoS One*, **10**, e0127446.
56. Serfiotis-Mitsa, D., Roberts, G.A., Cooper, L.P., White, J.H., Nutley, M., Cooper, A., Blakely, G.W. and Dryden, D.T.F. (2008) The Orf18 gene product from conjugative transposon Tn916 is an ArdA antirestriction protein that inhibits Type I DNA Restriction-Modification systems. *J. Mol. Biol.*, **383**, 970–981.
57. McMahon, S.A., Roberts, G.A., Johnson, K.A., Cooper, L.P., Liu, H., White, J.H., Carter, L.G., Sanghi, B., Oke, M., Walkinshaw, M.D. et al. (2009) Extensive DNA mimicry by the ArdA anti-restriction protein and its role in the spread of antibiotic resistance. *Nucleic Acids Res.*, **37**, 4887–4897.
58. Ellrott, K., Kasarjian, J., Jiang, T. and Ryu, J. (2002) Restriction enzyme recognition sequence search program. *Biotechniques*, **33**, 1322–1326.
59. Csefalvay, E., Lapkouski, M., Guzanova, A., Csefalvay, L., Baikova, T., Shevelev, I., Bialevich, V., Shamayeva, K., Janscak, P., Smatanova, I.K. et al. (2015) Functional coupling of duplex translocation to DNA cleavage in a type I restriction enzyme. *PLoS One*, **10**, e0128700.
60. Lapkouski, M., Panjikar, S., Janscak, P., Smatanova, I.K., Carey, J., Ettrich, R. and Csefalvay, E. (2009) Structure of the motor subunit of type I restriction-modification complex EcoR124I. *Nat. Struct. Mol. Biol.*, **16**, 94–95.
61. Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. and Sternberg, M.J. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.*, **10**, 845–858.
62. Powell, L.M., Lejeune, E., Hussain, F.S., Cronshaw, A.D., Kelly, S.M., Price, N.C. and Dryden, D.T. (2003) Assembly of EcoKI DNA methyltransferase requires the C-terminal region of the HsdM modification subunit. *Biophys. Chem.*, **103**, 129–137.
63. Powell, L.M., Dryden, D.T.F., Willcock, D.F., Pain, R.H. and Murray, N.E. (1993) DNA recognition by the EcoK methyltransferase. The influence of DNA methylation and the cofactor S-adenosyl-L-methionine. *J. Mol. Biol.*, **234**, 60–71.
64. Taylor, I.A., Davis, K.G., Watts, D. and Kneale, G.G. (1994) DNA-binding induces a major structural transition in a type I methyltransferase. *EMBO J.*, **13**, 5772–5778.
65. Dryden, D.T.F., Cooper, L.P., Thorpe, P.H. and Byron, O. (1997) The in vitro assembly of the EcoKI type I DNA restriction/modification enzyme and its in vivo implications. *Biochemistry*, **36**, 1065–1076.
66. Callow, P., Sukhodub, A., Taylor, J.E. and Kneale, G.G. (2007) Shape and subunit organisation of the DNA methyltransferase M.AhdI by small-angle neutron scattering. *J. Mol. Biol.*, **369**, 177–185.
67. Taylor, J.E., Callow, P., Swiderska, A. and Kneale, G.G. (2010) Structural and functional analysis of the engineered type I DNA methyltransferase EcoR124I(NT). *J. Mol. Biol.*, **398**, 391–399.
68. Liu, Y.P., Tang, Q., Zhang, J.Z., Tian, L.F., Gao, P. and Yan, X.X. (2017) Structural basis underlying complex assembly and conformational transition of the type I R-M system. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 11151–11156.
69. Dryden, D.T.F. (2017) Structures of the type I DNA restriction enzymes. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, E10261–E10262.
70. Dryden, D.T.F. (1999) Bacterial DNA Methyltransferases. In: *S-Adenosyl-methionine-dependent Methyltransferases: Structures and Functions*. World Scientific Publishing, pp. 283–340.
71. Koonin, E.V. and Martin, W. (2005) On the origin of genomes and cells within inorganic compartments. *Trends Genet.*, **12**, 647–654.
72. Koonin, E.V. and Wolf, Y.I. (2012) Evolution of microbes and viruses: a paradigm shift in evolutionary biology? *Front. Cell Infect. Microbiol.*, **2**, 119.
73. Koonin, E.V. and Dolja, V.V. (2013) A virocentric perspective on the evolution of life. *Curr. Opin. Virol.*, **5**, 546–557.
74. Weiss, M.C., Sousa, F.L., Mrnjavac, N., Neukirchen, S., Roettger, M., Nelson-Sathi, S. and Martin, W.F. (2016) The physiology and habitat of the last universal common ancestor. *Nat. Microbiol.*, **1**, 16116.
75. Singleton, M.R., Dillingham, M.S. and Wigley, D.B. (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annu. Rev. Biochem.*, **76**, 23–50.
76. Laurino, P., Tóth-Petróczy, Á., Meana-Pañeda, R., Lin, W., Truhlar, D.G. and Tawfik, D.S. (2016) An ancient fingerprint indicates the common ancestry of Rossmann-Fold enzymes utilizing different Ribose-Based cofactors. *PLoS Biol.*, **14**, e1002396.
77. Laurino, P. and Tawfik, D.S. (2017) Spontaneous Emergence of S-Adenosylmethionine and the Evolution of Methylation. *Angew. Chem. Int. Ed. Engl.*, **56**, 343–345.
78. Chang, C.C., Lin, L.Y., Zou, X.W., Huang, C.C. and Chan, N.L. (2015) Structural basis of the mercury(II)-mediated conformational switching of the dual-function transcriptional regulator MerR. *Nucleic Acids Res.*, **43**, 7612–7623.
79. Krüger, D.H. and Bickle, T.A. (1983) Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. *Microbiol. Rev.*, **47**, 345–360.
80. Tock, M.R. and Dryden, D.T.F. (2005) The biology of restriction and anti-restriction. *Curr. Opin. Microbiol.*, **8**, 466–472.
81. Dryden, D.T.F. (2006) DNA mimicry by proteins and the control of enzymatic activity on DNA. *Trends Biotechnol.*, **24**, 378–382.
82. Krüger, D.H., Reuter, M., Schroeder, C., Glatman, L.I. and Chernin, L.S. (1983) Restriction of bacteriophage T3 and T7 ocr+ strains by the type II restriction endonuclease EcoRV. *Mol. Gen. Genet.*, **190**, 349–351.
83. Weigle, P. and Raleigh, E.A. (2016) Biosynthesis and function of modified bases in bacteria and their viruses. *Chem. Rev.*, **116**, 12655–12687.

84. Jeltsch, A. (1999) Circular permutations in the molecular evolution of DNA methyltransferases. *J. Mol. Evol.*, **49**, 161–164.
85. Athanasiadis, A., Vlassi, M., Kotsifaki, D., Tucker, P.A., Wilson, K.S. and Kokkinidis, M. (1994) Crystal structure of PvuII endonuclease reveals extensive structural homologies to EcoRV. *Nat. Struct. Biol.*, **1**, 469–475.
86. Gong, W., O’Gara, M., Blumenthal, R.M. and Cheng, X. (1997) Structure of PvuII DNA-(cytosine N4) methyltransferase, an example of domain permutation and protein fold assignment. *Nucleic Acids Res.*, **25**, 2702–2715.
87. Halford, S.E. and Marko, J.F. (2004) How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.*, **32**, 3040–3052.
88. Marshall, J.J.T., Gowers, D.M. and Halford, S.E. (2007) Restriction endonucleases that bridge and excise two recognition sites from DNA. *J. Mol. Biol.*, **367**, 419–431.
89. Smith, R.M., Marshall, J.J., Jacklin, A.J., Retter, S.E., Halford, S.E. and Sobott, F. (2013) Organization of the BcgI restriction-modification protein for the cleavage of eight phosphodiester bonds in DNA. *Nucleic Acids Res.*, **41**, 391–404.
90. Gilmore, J.L., Suzuki, Y., Tamulaitis, G., Siksnys, V., Takeyasu, K. and Lyubchenko, Y.L. (2009) Single-molecule dynamics of the DNA-EcoRII protein complexes revealed with high-speed atomic force microscopy. *Biochemistry*, **48**, 10492–10498.
91. Shah, S., Sanchez, J., Stewart, A., Piperakis, M.M., Cosstick, R., Nichols, C., Park, C.K., Ma, X., Wysocki, V., Bitinaite, J. *et al.* (2015) Probing the run-on oligomer of activated SgrAI bound to DNA. *PLoS One*, **10**, e0124783.
92. Little, E.J., Babic, A.C. and Horton, N.C. (2008) Early interrogation and recognition of DNA sequence by indirect readout. *Structure*, **16**, 1828–1837.
93. Morgan, R.D., Luyten, Y.A., Johnson, S.A., Clough, E.M., Clark, T.A. and Roberts, R.J. (2016) Novel m4C modification in type I restriction-modification systems. *Nucleic Acids Res.*, **44**, 9413–9425.
94. Jeltsch, A. (2001) The cytosine N4-methyltransferase M.PvuII also modifies adenine residues. *Biol. Chem.*, **382**, 707–710.
95. Lauster, R. (1989) Evolution of type II DNA methyltransferases. A gene duplication model. *J. Mol. Biol.*, **206**, 313–321.
96. Mokrishcheva, M.L., Solonin, A.S. and Nikitin, D.V. (2011) Fused eco29kIR- and M genes coding for a fully functional hybrid polypeptide as a model of molecular evolution of restriction-modification systems. *BMC Evol. Biol.*, **11**, 35.
97. Walkinshaw, M.D., Taylor, P., Sturrock, S.S., Atanasiu, C., Berge, T., Henderson, R.M., Edwardson, J.M. and Dryden, D.T.F. (2002) Structure of Ocr from bacteriophage T7, a protein that mimics B-form DNA. *Mol. Cell*, **9**, 187–194.
98. Kanwar, N., Roberts, G.A., Cooper, L.P., Stephanou, A.S. and Dryden, D.T.F. (2016) The evolutionary pathway from a biologically inactive polypeptide sequence to a folded, active structural mimic of DNA. *Nucleic Acids Res.*, **44**, 4289–4303.
99. Roberts, G.A., Chen, K., Bower, E.K.M., Madrzak, J., Woods, A., Barker, A.M., Cooper, L.P., White, J.H., Blakely, G.W., Manfield, I. *et al.* (2013) Mutations of the domain forming the dimeric interface of the ArdA protein affect dimerization and antimodification activity but not antirestriction activity. *FEBS J.*, **280**, 4903–4914.
100. Spoerel, N., Herrlich, P. and Bickle, T.A. (1979) A novel bacteriophage defence mechanism: the anti-restriction protein. *Nature*, **278**, 30–34.
101. Loenen, W.A.M. and Murray, N.E. (1986) Modification enhancement by the restriction alleviation protein (Ral) of bacteriophage lambda. *J. Mol. Biol.*, **190**, 11–22.
102. King, G. and Murray, N.E. (1995) Restriction alleviation and modification enhancement by the Rac prophage of *Escherichia coli* K-12. *Mol. Microbiol.*, **16**, 769–777.
103. Neaves, K.J., Cooper, L.P., White, J.H., Carnally, S.M., Dryden, D.T.F., Edwardson, J.M. and Henderson, R.M. (2009) Atomic force microscopy of the EcoKI Type I DNA restriction enzyme bound to DNA shows enzyme dimerization and DNA looping. *Nucleic Acids Res.*, **37**, 2053–2063.
104. Crampton, N., Roes, S., Dryden, D.T.F., Rao, D.N., Edwardson, J.M. and Henderson, R.M. (2007) DNA looping and translocation provide an optimal cleavage mechanism for the type III restriction enzymes. *EMBO J.*, **26**, 3815–3825.
105. Wu, D., Kaur, P., Li, Z.M., Bradford, K.C., Wang, H. and Erie, D.A. (2016) Visualizing the path of DNA through proteins using DREEM imaging. *Mol. Cell*, **61**, 315–323.